Medical Text Indexer (MTI)

Title + Abstract

Phrasex

Noun Phrases

MetaMap

Trigram

PubMed Related Citations

UMLS concepts

Rel. Citations

Restrict to MeSH

Extract MeSH descr.

MeSH Main Headings

Clustering & Ranking

Ordered list of MeSH Main Headings

(Last Updated: January 8, 2003)

Index

List of Equations

List of Figures

1. <u>Introduction</u>

This document will provide detailed information on the behind the scenes processing that takes place in the Medical Text Indexer (MTI). As the diagram below shows, there is a lot going on.
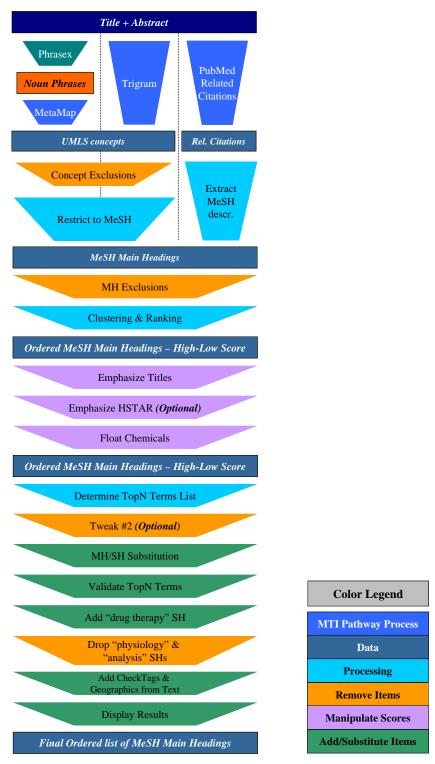

Figure 1: Detailed Medical Text Indexer Process Flow Diagram

2. Exclusions

The following terms are excluded or substituted for even before we get to the Clustering phase of processing in the MTI system.

Regardless of pathway (MetaMap or PubMed) the following MeSH terms are removed:
- TEST
- Comparative Study
- Case Report
- Disease
- [Publication Type]

If the pathway is MetaMap, the following MeSH term is removed:
- Role is excluded

If the pathway is MetaMap ONLY, the following MeSH term replacements are done:
- Men is replaced by CheckTag Male and CheckTag Human

- Women is replaced by CheckTag Female and CheckTag Human

- Patients is replaced by CheckTag Human

If the **remMHs** parameter is set:

- Remove MeSH Headings found in the MH Exclusion list (*Appendix-G*) regardless of path.

3. Clustering and Ranking

The MeSH® headings produced by all of the Indexing Initiative (II) methods are clustered and then formed into a single, final list of recommended indexing terms. This document discusses the steps involved in this clustering and ranking process. A high level view of the steps involved in the processing is as follows:

1. Load and summarize individual path results calculating the term weights (see the section entitled "*Calculating Term Weights*"),

2. Clustering of the results – determining which of the results are related (see the section entitled "*Clustering*"), and finally

3. Ranking the results – using the information obtained in 1 and 2 to compute the rank of each item (see the section entitled "*Ranking*").

Each of these steps will be reviewed in detail over the following sections of this document. But first we provide the reader with some background on where the underlying data used in the processing comes from in the next two sections.

3.1. Overview of Clustering and Ranking (from BoSC99 report)

The task here is to provide a weighting of the confidence or strength of belief in the assignment, and rank the suggested headings appropriately. There are a number of factors that can be recognized as playing a role in that confidence. The method of finding the heading (the path), how much confidence is available in how the method found the heading (the goodness of the match), the location in the text of the nominal phrase that led to that suggestion (the location), and the semantic consistency of the suggested heading with the other suggested headings (the corroborating evidence).

Assigning a weight to the overall method of finding the heading (the *PathWeight*) allows one to discount a method appropriate to strengths. For example, a certain path might not be very specific, but have some sensitivity in suggesting headings that would otherwise not occur. When headings found by other paths offer corroborative evidence for a heading suggested by this method, the additional confidence gained might be helpful.

The goodness of the match, i.e., how much confidence to place in a given heading, depends on the method used to find the heading. The possibilities are:

- A phrase identified in text is an exact match to a MeSH term. Equivalently, it might have been a match to a UMLS® term that was a synonym of a MeSH term.

- Of lesser significance is an exact match to a UMLS term that is then be mapped to a MeSH heading using the Restrict to MeSH method.

- Another possibility is that the phrase is an inexact, or approximate, match to a UMLS term, which is either a synonym of a MeSH heading or mapped to MeSH.

Thus, each time a MeSH heading is suggested, a weighting can be given to that suggestion. This is accomplished using both a *MapScore* and a *NavScore*. The MapScore reflects the confidence in the mapping to a UMLS term, the NavScore the confidence in navigating from a UMLS term to a MeSH Heading.

With regard to the importance of location, the main consideration was whether or not the phrase leading to a heading suggestion was mentioned in the title. All other things being equal, indexers know that things mentioned in the title of the article are probably more important than other concepts mentioned in the article. Similarly, if the heading was suggested by a phrase occurring in the title, it should be given more weight. The additional weight is added as a constant in the formula.

Semantic consistency can be thought of as corroborative evidence for the goodness of a suggestion. It is identified by relationships that a suggested heading has with other suggested headings. These relationships might be either the occurrence in the same hierarchy (as parents or siblings), or as known co-occurring headings in MEDLINE. This latter evidence needs to be weighted according to a normalized frequency of this co-occurrence. The normalized frequency times a constant becomes the COT weight. The former evidence is the REL weight, and is a simple constant.

The overall RankScore can be altered by changing any of the constants (COT, REL, and PathWeight) or by changing the method by which the weight is calculated (NavScore and MapScore). Altering these values allows a number of experiments to be performed to evaluate the robustness of the weighting scheme, and to establish reasonable values for the constants.

3.2. UMLS® Metathesaurus® Files

There are two main UMLS Metathesaurus files used by the clustering and ranking functions, the MRREL and MRCOC files. The following definitions come directly from the UMLS Metathesaurus documentation. The MRCOC file is used to create the normalized frequency database table that the Indexing Initiative uses.

3.2.1. Related Concepts (File = MRREL)

There is one row in this table for each relationship between Metathesaurus concepts known to the Metathesaurus, with the following exceptions found in other files: co-occurrences found in MRCOC; Locator information in MRLO; and Associated Expressions found in MRATX.

Note that for asymmetrical relationships there is one row for each direction of the relationship. Note also the direction of REL - the relationship that the SECOND concept (with Concept Unique Identifier CUI2) HAS TO the FIRST concept (with Concept Unique Identifier CUI1).

RELs may be derived from a source vocabulary's explicit hierarchy (see also MRCXT), derived from other relationships in a source vocabulary, created from information about allowed qualifiers in a source vocabulary, found in Metathesaurus QA of lexical and semantic matches, or added by Metathesaurus editors.

Where relationships are asymmetrical, there are separate RELS for each direction of the relationship, e.g., one entry for "Atrial Fibrillation" as a child of "Arrhythmia" and another entry for "Arrhythmia" as a parent of "Atrial Fibrillation".

Valid Values for REL:

RB	has a broader relationship
RN	has a narrower relationship
RO	has relationship other than synonymous, narrower, or broader
RL	the relationship is similar or "alike". Some concepts linked by the RL relationship may be determined to be synonyms in future editions of the Metathesaurus. In the current edition of the Metathesaurus, most RL relationships link MeSH supplementary concepts, which have not yet been edited in the new MeSH concept-oriented system. In future editions of the Metathesaurus, this Relation will also be used for "quasisynonyms", such as "Hypertension" and "High Blood Pressure", which are sometimes used synonymously, but have distinct meanings in some circumstances. When RL is used for quasisynonyms, the RELA (Relationship Attribute) will further identify the "quasisynonymous" Relationship.
PAR	has parent relationship in a Metathesaurus source vocabulary
CHD	has child relationship in a Metathesaurus source vocabulary
SIB	has sibling relationship in a Metathesaurus source vocabulary.
AQ	is an allowed qualifier for the first concept in a Metathesaurus source vocabulary.

3.2.2. Co-occurring Concepts (File = MRCOC)

There are two rows in this table for each pair of concepts that co-occur in each information source represented one for each direction of the relationship. (Note that the COA data may be different for each direction of the relationship). Many Metathesaurus concepts have no entries in this file. Due to the very large number of co-occurrence relationships, they are distributed in a separate file.

Co-occurrences are concepts that occur together in the same "entries" in some information source. The relationships represented here are obtained from machine-manipulation of the information source. Co-occurrence relationships may exist between similar concepts (e.g., "Atrial Fibrillation" and "Arrhythmia") or between very different concepts that nevertheless have some important connection in the field of biomedicine (e.g., "Atrial Fibrillation" and "Digoxin"), or between a primary concept and a qualifier e.g., "Lithotripsy" and "instrumentation". A co-occurrence relationship can exist between two concepts that have no other apparent relationship, although the frequency of such co-occurrences will be small.

In the current Metathesaurus, there are three sources of co-occurrence data: MEDLINE, AI/RHEUM, and CCPSS. From MEDLINE, co-occurrence data was computed for concepts that were designated as principal or main points in the same journal article i.e., the co-occurrence counts do not include articles in which either or both of the concepts were present and indexed in MEDLINE but not designated as main points. (A concept is considered to be a main point if the * is attached to the main heading or any of its subheadings.)

3.3. Creating the Normalized Frequency Scores for the Co-Occurring Concepts

This section of the document discusses how we create the co-occurring concepts normalized frequency database used in the Indexing Initiative's Medical Text Indexer (MTI). These steps are done once at the beginning of each year with the final released version of the UMLS Metathesaurus, specifically the MRCOC table.

3.3.1. Overview
The following steps calculate the normalized frequency score for the co-occurring concepts:

1. Summarize all of the records we have by combining identical pairings of CUI1 and CUI2 frequency counts,

2. Determine an overall total of frequency counts for each CUI1 we have, and

3. Finally, divide the frequency counts for each of the records (now summarized) by the total number of frequency counts for the CUI1 that the record is associated with.

3.3.2. Detailed Explanation and Example

1. We pull all records from the MRCOC file except ones containing "|LQ|" in the Type of Co-Occurrence (COT) field. The "LQ" (MeSH topical qualifier) records are only relevant if we want to augment our SubHeading recommendations. We only keep fields 1, 2, and 5 -- CUI1, CUI2, and COF (Frequency of Co-Occurrence) respectively in a bar separated list.

   We end up with a file containing lines similar to the sample below:

   ```
   C0000039|C0000300|2
   C0000039|C0001006|1
   C0000039|C0001128|1
   C0000039|C0001392|1
   C0000039|C0001480|1
   C0000039|C0001480|1
   …
   ```

2. We than summarize this list by CUI1 by summing the COF for each CUI1 and CUI2 combination and providing a total frequency count for each CUI1 and CUI2 pairing. In the example in #1 above, we would combine the last two rows because the CUI1 and CUI2 pairings are identical. We end up with a file containing lines similar to the sample below:

   ```
   C0000039|C0000300|2
   C0000039|C0001006|1
   C0000039|C0001128|1
   C0000039|C0001392|1
   C0000039|C0001480|2
   …
   ```

3. We create a temporary file containing a single line for each unique CUI1 concept. This line contains the total frequency count for that particular CUI1. We end up with a file containing lines similar to the sample below:

   ```
   C0000039|1190
   …
   ```

4. We now combine the two files from #2 and #3. We want to end up with a file containing all of the records of #2 above and the total frequency count from #3 above appended to the end of the line. We end up with a file containing lines similar to the sample below:

   ```
   C0000039|C0000300|2|1190
   C0000039|C0001006|1|1190
   C0000039|C0001128|1|1190
   C0000039|C0001392|1|1190
   C0000039|C0001480|2|1190
   …
   ```

5. We now calculate the normalization of the frequency counts for each of the records by dividing the individual record's frequency count (field 3) by the CUI1's total frequency count (field 4). We end up with a file containing lines similar to the sample below:

```
C0000039|C0000300|0.001681
C0000039|C0001006|0.000840
C0000039|C0001128|0.000840
C0000039|C0001392|0.000840
C0000039|C0001480|0.001681
…
```

3.4. Calculating TermWeight

The TermWeight for each MeSH Heading is the summation of all entries for a MH from each of the various paths used (MetaMap after Restrict to MeSH (MMI) and PubMed Related Citations (RC)). The TermWeight for each MH regardless of path is calculated using the following formula where *i* represents the single occurrence of the suggestion of one MeSH heading:

$$TermWeight = TW = \sum_{i=1}^{n} (PathWeight_i * MapScore_i * NavScore_i)$$

Equation 1 - TermWeight Formula

Assigning a weight to the overall method of finding the heading (*PathWeight*) allows one to discount a method appropriate to strengths. The *MapScore* reflects the confidence in the mapping to an UMLS term by a specific method, the *NavScore* is the confidence in navigating from an UMLS term to a MeSH heading.

3.4.1. Tunable and System Parameters

The following table depicts the parameters used in calculating the TermWeight along with their default values:

Abbreviation	Full Name	Tunable by User	Default Value
MMI	MetaMap Indexing Path Weight (*PathWeight*)	X	7
RC	Related Citations Path Weight (*PathWeight*)	X	2
I	Direct Match Navigational String – MMI (*NavScore*)	X	1.00
A	ATX (Associated Expression) Navigational String – MMI (*NavScore*)	X	1.00
G/P	Parent/Broader Navigational String – MMI (*NavScore*)	X	0.90
G/C	Child/Narrower Navigational String – MMI (*NavScore*)	X	0.75
G/S	Sibling Navigational String – MMI (*NavScore*)	X	0.70
O	Other Related Navigational String – MMI (*NavScore*)	X	0.50
IM	MeSH Major Topic Navigational String – RC (*NavScore*)	X	1.00
NIM	MeSH Heading Navigational String – RC (*NavScore*)	X	0.80
Best possible score for items returned by the MMI path (*MapScore*)		-	1,000
Best possible score for items returned by the RC path (*MapScore*)		-	255
Best possible score for items returned by the Trigram path (*MapScore*)		-	1,000

3.4.2. Steps Followed in Calculating the TermWeight

The following steps are done for each MeSH Heading:

1. The weight from the item is provided by each of the individual paths along with the navigational string information. The following example shows items returned for the concept "Blood Flow Velocity" via both the MMI and RC pathways. The individual MapScores are highlighted in **blue** and the individual navigation strings are highlighted in **tan**.

   ```
   ❶ MMI: 97479605|C0005798|118|G/P|Blood Flow Velocity|MH||TI|
   ❷ MMI: 97479605|C0005798|118|O|Blood Flow Velocity|MH||TI|
   ❸ RC:  97479605|C0005798|28.1847|NIM|Blood Flow Velocity|MH|3|
   ❹ RC:  97479605|C0005798|26.4019|NIM|Blood Flow Velocity|MH|8|
   ❺ RC:  97479605|C0005798|26.0665|NIM|Blood Flow Velocity|MH|10|
   ```

   In the first line we have an item coming from the MMI pathway with a MapScore of 118 out of a possible 1,000 perfect score and having a navigational string of G/P (Parent/Broader).

   In the third line we have an item coming from the RC pathway with a MapScore of 28.1847 out of a possible 255 perfect score and having a navigational string of NIM (MeSH Heading).

2. The items are loaded into the program systematically, so we will always load all of the MMI terms before loading all of the RC terms. To calculate the PathWeight to be used in the calculations for each item, we divide the individual path weight by the path-scoring factor. The path-scoring factor is used to equalize all of the different scoring methods. If the path is MMI or Trigram, we use 1,000 and for RC, we use 255.

   ```
   MMI PathWeight = 7/1000 = 0.0070
   RC PathWeight = 2/255 = 0.0078
   ```

3. We can than calculate the individual item weights via (PathWeight * MapScore * NavScore) where NavScore depends on the navigation string (see table above):

   ```
   ❶ MMI: (118 * 0.0070) * 0.90 (G/P) = 0.7434
   ❷ MMI: (118 * 0.0070) * 0.50 (O) = 0.4130
   ❸ RC:  (28.1847 * 0.0078) * 0.80 (NIM) = 0.1769
   ❹ RC:  (26.4019 * 0.0078) * 0.80 (NIM) = 0.1657
   ❺ RC:  (26.0665 * 0.0078) * 0.80 (NIM) = 0.1635
   ```

4. Now we sum all of these individual item weights together to get our final TermWeight.

$$0.7434\text{❶} + 0.4130\text{❷} + 0.1769\text{❸} + 0.1657\text{❹} + 0.1635\text{❺} = 1.6625$$

For our example "Blood Flow Velocity", we have a final TermWeight of 1.6625 and the five (5) different path entries have been summarized into a single term in our list containing the concept name, CUI, score (which is zero at this point), and the TermWeight that we just calculated.

Blood Flow Velocity|C0005798|0|1.6625

5. The summarized list for all processed and summarized items will look similar to the following:

```
mt_table[0]:  DNA-Binding Proteins|C0012940|0|1.0150
mt_table[1]:  Transcription Factors|C0040648|0|1.0150
mt_table[2]:  SEF1 protein|C0212321|0|1.0150
mt_table[3]:  Blood Circulation Time|C0919393|0|1.1564
mt_table[4]:  Radionuclide Imaging|C0034606|0|1.1564
mt_table[5]:  Blood Flow Velocity|C0005798|0|1.6625
mt_table[6]:  Neurology|C0027855|0|0.4025
. . .
mt_table[84]:  Confusion|C0009676|0|0.1651
mt_table[85]:  Glasgow Coma Scale|C0017594|0|0.3287
mt_table[86]:  Predictive Value of Tests|C0032944|0|0.1651
mt_table[87]:  Regional Blood Flow|C0034965|0|0.1651
mt_table[88]:  Regression Analysis|C0034980|0|0.1651
```

3.5. Clustering

In the clustering phase, we are going to go through every item in our summarized and term weighted list looking for what other items in the list either co-occur with the item or are related via the MeSH tree structure to the item. In an attempt to make the process faster, we are going to compute the clustering in both directions as we progress through the items list. This means we only have to make a single pass through the list. The table below depicts how we progress through the item list computing from the item we are currently working on forward to the end of the item list. This works because the co-occurring and MeSH tree relationship lists should always be symmetrical (e.g., if we have an entry A|B we also have an entry B|A) as defined by the UMLS Metathesaurus (see section entitled "UMLS Metathesaurus Files").

	0	1	2	. . .	86	87	88
0							
1	*						
2	*	*					
. . .	*	*	*				
86	*	*	*	*			
87	*	*	*	*	*		
88	*	*	*	*	*	*	

Figure 2: Picture of how we traverse the item list for clustering

The results of the clustering process are compartmented into co-occurring terms (cot) and MeSH tree relationship terms. The MeSH tree relationships are again compartmented into PAR/CHD/SIB (treerel) and then RN/RB/RO (othrel) (see section entitled "Related Concepts (File = MRREL)" for definitions).

3.5.1. Overview of Steps for Clustering

- For every item (i) in our summarized and term weighted list we do the following:

  1. For every item remaining (k) in our list ahead of i (e.g., i + 1 to n), we do the following:

     1) Retrieve the CUIs for item[i] and item[k]

     2) See if we have a co-occurring match of the item[i] and item[k] CUIs. If we do,

        i. Add an entry into item[i]'s cot list containing item[k]'s concept name, normalized frequency, and TermWeight.

        ii. Verify that we have the symmetrical co-occurring match of the item[k] and item[i] CUIs and add an entry into item[k]'s cot list containing item[i]'s concept name, normalized frequency, and TermWeight. **Note:** *We might not have a match since we have removed some of the really low normalized frequency count items.*

     3) See if we have a MeSH tree relationship match of the item[i] and item[k] CUIs. If we do, then for each match we have (**Note:** *there can be multiple MeSH tree relationship results*) do the following:

        i. Retrieve the relationship information from the match result and then:

        ii. If the relationship is Parent, Child, or Sibling (PAR/CHD/SIB) then:

           1. Add an entry into item[i]'s treerel list containing item[k]'s concept name, normalized frequency, and TermWeight.

           2. Add the symmetrical entry by adding an entry into item[k]'s treerel list containing item[i]'s concept name, normalized frequency, and TermWeight.

        iii. If the relationship is Broader, Narrower, or Other (RN/RB/RO) then:

           1. Add an entry into item[i]'s othrel list containing item[k]'s concept name, normalized frequency, and TermWeight.

           2. Add the symmetrical entry by adding an entry into item[k]'s othrel list containing item[i]'s concept name, normalized frequency, and TermWeight.

3.5.2. Example of Clustering

In this example, we continue using our example concept "Blood Flow Velocity". Here we are showing the effects of the clustering on our concept. We have tied each of the example steps below to the steps described in the overview above by adding notations at the beginning of each line (e.g., (2.i) means step 2.i as described in the overview section). The first entry in the example below (mt_table[3] …) which is highlighted and annotated provides us with a good look at all of the aspects of the clustering process. "Blood Circulation Time"

1) We have an item that does not co-occur with and doesn't relate to in "Radionuclide Imaging",

2) We have an item that does co-occur with in "Blood Flow Velocity",

3) The "Blood Flow Velocity" item is not symmetrical since the inverse pairing was removed from the co-occurring table due to it's small normalized frequency count ("NOT FOUND"),

4) The "Blood Flow Velocity" item is related via the MeSH tree structure as a relationship other than synonymous, narrower, or broader (RO),

5) The "Blood Flow Velocity" item is related via the MeSH tree structure as a sibling (SIB),

6) Finally, the "Blood Flow Velocity" item shows how we are handling both directions in the single pass of clustering. The symmetrical entry for the "Blood Flow Velocity" item is automatically added here and doesn't need to be reviewed when we get to it later on.

```
(i) mt_table[3]: Blood Circulation Time
   (k) mt_table[4]: Radionuclide Imaging        ←----\
      (2) No Co-Occurring Terms Found           ←---------- (1)
      (3) No MeSH Tree Related Terms Found      ←----/

   (k) mt_table[5]: Blood Flow Velocity
      (2.i) Co-Occurring Normalized Frequency: 0.0556  ←------- (2)
      (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: NOT FOUND  ←----- (3)
      (3.i) MeSH Tree Relationship: RO    ←---------------- (4)
      (3.i) MeSH Tree Relationship: SIB   ←---------------- (5)
                . . .
(i) mt_table[5]: Blood Flow Velocity
   (k) mt_table[6]: Neurology
      (2) No Co-Occurring Terms Found
      (3) No MeSH Tree Related Terms Found
                . . .
   (k) mt_table[10]: Infant, Newborn
      (2.i) Co-Occurring Normalized Frequency: 0.0014
      (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0007
      (3) No MeSH Tree Related Terms Found
                . . .
   (k) mt_table[19]: Tomography, Emission-Computed
      (2.i) Co-Occurring Normalized Frequency: 0.0007
      (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: NOT FOUND
      (3) No MeSH Tree Related Terms Found
                . . .
   (k) mt_table[29]: Brain
      (2.i) Co-Occurring Normalized Frequency: 0.0072
      (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: NOT FOUND
      (3) No MeSH Tree Related Terms Found
```

```
                . . .
        (k) mt_table[33]: Homeostasis
            (2.i) Co-Occurring Normalized Frequency: 0.0012
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0007
            (3) No MeSH Tree Related Terms Found
                . . .
        (k) mt_table[35]: Craniocerebral Trauma
            (2.i) Co-Occurring Normalized Frequency: 0.0010
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0006
            (3) No MeSH Tree Related Terms Found
                . . .
        (k) mt_table[40]: Blood Vessels
            (2.i) Co-Occurring Normalized Frequency: 0.0067
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0048
            (3) No MeSH Tree Related Terms Found

        (k) mt_table[41]: Vascular Diseases
            (2.i) Co-Occurring Normalized Frequency: 0.0012
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0009
            (3) No MeSH Tree Related Terms Found
                . . .
        (k) mt_table[46]: Cerebrovascular Circulation
            (2.i) Co-Occurring Normalized Frequency: 0.0244
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0046
            (3) No MeSH Tree Related Terms Found

        (k) mt_table[47]: Gestational Age
            (2.i) Co-Occurring Normalized Frequency: 0.0005
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0005
            (3) No MeSH Tree Related Terms Found

        (k) mt_table[48]: Infant, Premature
            (2.i) Co-Occurring Normalized Frequency: 0.0017
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0005
            (3) No MeSH Tree Related Terms Found
                . . .
        (k) mt_table[50]: Brain Ischemia
            (2.i) Co-Occurring Normalized Frequency: 0.0012
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: NOT FOUND
            (3) No MeSH Tree Related Terms Found
                . . .
        (k) mt_table[52]: Intracranial Pressure
            (2.i) Co-Occurring Normalized Frequency: 0.0012
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0014
            (3) No MeSH Tree Related Terms Found

        (k) mt_table[53]: Oxygen
            (2.i) Co-Occurring Normalized Frequency: 0.0043
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0005
            (3) No MeSH Tree Related Terms Found
                . . .
        (k) mt_table[55]: Heart Rate
            (2.i) Co-Occurring Normalized Frequency: 0.0017
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: NOT FOUND
            (3.i) MeSH Tree Relationship: SIB
                . . .
        (k) mt_table[61]: Umbilical Arteries
            (2.i) Co-Occurring Normalized Frequency: 0.0112
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0163
            (3) No MeSH Tree Related Terms Found

        (k) mt_table[62]: Xenon Radioisotopes
            (2.i) Co-Occurring Normalized Frequency: 0.0005
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0041
            (3) No MeSH Tree Related Terms Found
                . . .
        (k) mt_table[64]: Cerebrospinal Fluid Pressure
            (2.i) Co-Occurring Normalized Frequency: 0.0005
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0027
            (3) No MeSH Tree Related Terms Found
                . . .
```

```
        (k) mt_table[70]: Tomography, X-Ray Computed
            (2.i) Co-Occurring Normalized Frequency: 0.0007
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: NOT FOUND
            (3) No MeSH Tree Related Terms Found

        (k) mt_table[71]: Blood Pressure
            (2.i) Co-Occurring Normalized Frequency: 0.0184
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0017
            (3.i) MeSH Tree Relationship: SIB
                    . . .
        (k) mt_table[74]: Aging
            (2.i) Co-Occurring Normalized Frequency: 0.0031
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: NOT FOUND
            (3) No MeSH Tree Related Terms Found
                    . . .
        (k) mt_table[77]: Echoencephalography
            (2.i) Co-Occurring Normalized Frequency: 0.0024
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0061
            (3) No MeSH Tree Related Terms Found

        (k) mt_table[78]: Linear Models
            (2.i) Co-Occurring Normalized Frequency: 0.0005
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0018
            (3) No MeSH Tree Related Terms Found
                    . . .
        (k) mt_table[87]: Regional Blood Flow
            (2.i) Co-Occurring Normalized Frequency: 0.0014
            (2.ii) Co-Occurring Symmetrical Match Normalized Frequency: 0.0058
            (3.i) MeSH Tree Relationship: SIB

        (k) mt_table[88]: Regression Analysis
            (2) No Co-Occurring Terms Found
            (3) No MeSH Tree Related Terms Found
```

After all of the clustering process has completed – the following list contains the co-occurrence and relationship information we have accumulated for our example. This information consists of concept name, normalized frequency, and TermWeight for each item clustered together with our term.

```
mt_table[5]: Blood Flow Velocity|C0005798|0|1.6625
     cot[0]: Infant, Newborn  normfreq: 0.0014  termweight: 0.9240
     cot[1]: Tomography, Emission-Computed  normfreq: 0.0007  termweight: 0.4136
     cot[2]: Brain  normfreq: 0.0072  termweight: 1.1118
     cot[3]: Homeostasis  normfreq: 0.0012  termweight: 0.4327
     cot[4]: Craniocerebral Trauma  normfreq: 0.0010  termweight: 0.2316
     cot[5]: Blood Vessels  normfreq: 0.0067  termweight: 0.0105
     cot[6]: Vascular Diseases  normfreq: 0.0012  termweight: 0.0105
     cot[7]: Cerebrovascular Circulation  normfreq: 0.0244  termweight: 2.1469
     cot[8]: Gestational Age  normfreq: 0.0005  termweight: 0.3901
     cot[9]: Infant, Premature  normfreq: 0.0017  termweight: 0.4203
     cot[10]: Brain Ischemia  normfreq: 0.0012  termweight: 0.8588
     cot[11]: Intracranial Pressure  normfreq: 0.0012  termweight: 0.5694
     cot[12]: Oxygen  normfreq: 0.0043  termweight: 0.3911
     cot[13]: Heart Rate  normfreq: 0.0017  termweight: 0.1768
     cot[14]: Umbilical Arteries  normfreq: 0.0112  termweight: 0.1768
     cot[15]: Xenon Radioisotopes  normfreq: 0.0005  termweight: 0.3425
     cot[16]: Cerebrospinal Fluid Pressure  normfreq: 0.0005  termweight: 0.2144
     cot[17]: Tomography, X-Ray Computed  normfreq: 0.0007  termweight: 0.3767
     cot[18]: Blood Pressure  normfreq: 0.0184  termweight: 0.1659
     cot[19]: Aging  normfreq: 0.0031  termweight: 0.1657
     cot[20]: Echoencephalography  normfreq: 0.0024  termweight: 0.1657
     cot[21]: Linear Models  normfreq: 0.0005  termweight: 0.1657
     cot[22]: Regional Blood Flow  normfreq: 0.0014  termweight: 0.1651

     treerel[0]: Blood Circulation Time  rel: SIB  termweight: 1.1564
     treerel[1]: Heart Rate  rel: SIB  termweight: 0.1768
     treerel[2]: Blood Pressure  rel: SIB  termweight: 0.1659
     treerel[3]: Regional Blood Flow  rel: SIB  termweight: 0.1651

     othrel[0]: Blood Circulation Time  rel: RO  termweight: 1.1564
```

3.6. Calculating the RankScore

This is the final stage were we go through all of the information saved from the previous steps and calculate a final RankScore for each item based on the TermWeight, the normalized frequency count, and user specified constants for COT, REL, Title, and PathWeight. The formula for the RankScore is as follows:

$$RankScore = TW * \left[F * \left[1 + \sum_{j=1}^{n}(COT_j * TW_j) + \sum_{k=1}^{n}(REL * TW_k) \right] \right]$$

Equation 2 - RankScore Formula

The following table depicts the user-defined parameters we use in calculating the RankScore along with their default values:

Abbreviation	Description	Tunable by User	Default Value
COT	Factor for Co-Occurring Terms	X	10,000
REL	Factor for Tree Relationship	X	100
TW	TermWeight	-	-
F	Path Factor: If the item comes from MetaMap or Trigrams AND also from PubMed Related Citations F = 2 otherwise F = 1	-	-

3.6.1. Summary of Steps for Calculating the RankScore

1. Set score = 1,

2. Compute scores for the co-occurring terms,

3. Add scores for the PAR/CHD/SIB MeSH tree related terms,

4. Add scores for the RN/RB/RO MeSH tree related terms,

5. Set up and factor in the Path Factor for the item based on what paths recommended the item, and

6. Finally, factor in the TermWeight for the item into the score.

3.6.2. Example of Calculating the RankScore

1. Set score = 1

2. For each of the co-occurring terms (cot) items we found in clustering, do the following:

    score = score + (**term's normalized frequency count** * COT * **term's term weight**)

    ```
    cot[0]: Infant, Newborn  normfreq: 0.0014  termweight: 0.9240
      score = 1.0000 + (0.0014 * 10000 * 0.9240) == 14.2594
    cot[1]: Tomography, Emission-Computed  normfreq: 0.0007  termweight: 0.4136
      score = 14.2594 + (0.0007 * 10000 * 0.4136) == 17.2287
    cot[2]: Brain  normfreq: 0.0072  termweight: 1.1118
      score = 17.2287 + (0.0072 * 10000 * 1.1118) == 97.0027
    cot[3]: Homeostasis  normfreq: 0.0012  termweight: 0.4327
      score = 97.0027 + (0.0012 * 10000 * 0.4327) == 102.1775
    cot[4]: Craniocerebral Trauma  normfreq: 0.0010  termweight: 0.2316
      score 775 = 102.1775 + (0.0010 * 10000 * 0.2316) == 104.3940
    cot[5]: Blood Vessels  normfreq: 0.0067  termweight: 0.0105
      score = 104.3940 + (0.0067 * 10000 * 0.0105) == 105.0971
    cot[6]: Vascular Diseases  normfreq: 0.0012  termweight: 0.0105
      score = 105.0971 + (0.0012 * 10000 * 0.0105) == 105.2227
    cot[7]: Cerebrovascular Circulation  normfreq: 0.0244  termweight: 2.1469
      score = 105.2227 + (0.0244 * 10000 * 2.1469) == 628.9811
    cot[8]: Gestational Age  normfreq: 0.0005  termweight: 0.3901
      score = 628.9811 + (0.0005 * 10000 * 0.3901) == 630.8457
    cot[9]: Infant, Premature  normfreq: 0.0017  termweight: 0.4203
      score = 630.8457 + (0.0017 * 10000 * 0.4203) == 637.8818
    cot[10]: Brain Ischemia  normfreq: 0.0012  termweight: 0.8588
      score = 637.8818 + (0.0012 * 10000 * 0.8588) == 648.1532
    cot[11]: Intracranial Pressure  normfreq: 0.0012  termweight: 0.5694
      score = 648.1532 + (0.0012 * 10000 * 0.5694) == 654.9635
    cot[12]: Oxygen  normfreq: 0.0043  termweight: 0.3911
      score = 654.9635 + (0.0043 * 10000 * 0.3911) == 671.8002
    cot[13]: Heart Rate  normfreq: 0.0017  termweight: 0.1768
      score = 671.8002 + (0.0017 * 10000 * 0.1768) == 674.7606
    cot[14]: Umbilical Arteries  normfreq: 0.0112  termweight: 0.1768
      score = 674.7606 + (0.0112 * 10000 * 0.1768) == 694.6397
    cot[15]: Xenon Radioisotopes  normfreq: 0.0005  termweight: 0.3425
      score = 694.6397 + (0.0005 * 10000 * 0.3425) == 696.2769
    cot[16]: Cerebrospinal Fluid Pressure  normfreq: 0.0005  termweight: 0.2144
      score = 696.2769 + (0.0005 * 10000 * 0.2144) == 697.3019
    cot[17]: Tomography, X-Ray Computed  normfreq: 0.0007  termweight: 0.3767
      score = 697.3019 + (0.0007 * 10000 * 0.3767) == 700.0068
    cot[18]: Blood Pressure  normfreq: 0.0184  termweight: 0.1659
      score = 700.0068 + (0.0184 * 10000 * 0.1659) == 730.5548
    cot[19]: Aging  normfreq: 0.0031  termweight: 0.1657
      score = 730.5548 + (0.0031 * 10000 * 0.1657) == 735.7051
    cot[20]: Echoencephalography  normfreq: 0.0024  termweight: 0.1657
      score = 735.7051 + (0.0024 * 10000 * 0.1657) == 739.6677
    cot[21]: Linear Models  normfreq: 0.0005  termweight: 0.1657
      score = 739.6677 + (0.0005 * 10000 * 0.1657) == 740.4596
    cot[22]: Regional Blood Flow  normfreq: 0.0014  termweight: 0.1651
      score = 740.4596 + (0.0014 * 10000 * 0.1651) == 742.8291
    ```

The score at the end of processing the co-occurring terms is 742.8291.

3. For each of the PAR/CHD/SIB MeSH tree related terms (treerel) items we found in clustering, do the following:

score = score + (**term's term weight** * REL)

```
treerel[0]: Blood Circulation Time  rel: SIB  termweight: 1.1564
  score = 742.8291 + (1.1564 * 100) == 858.4691
treerel[1]: Heart Rate  rel: SIB  termweight: 0.1768
  score = 858.4691 + (0.1768 * 100) == 876.1536
treerel[2]: Blood Pressure  rel: SIB  termweight: 0.1659
  score = 876.1536 + (0.1659 * 100) == 892.7405
treerel[3]: Regional Blood Flow  rel: SIB  termweight: 0.1651
  score = 892.7405 + (0.1651 * 100) == 909.2529
```

The score at the end of processing the PAR/CHD/SIB MeSH tree related terms is 909.2529.

4. For each of the RN/RB/RO MeSH tree related terms (othrel) items we found in clustering, do the following:

score = score + (**term's term weight** * REL)

```
othrel[0]: Blood Circulation Time  rel: RO  termweight: 1.1564
  score = 909.2529 + (1.1564 * 100) == 1024.8929
```

The score at the end of processing the RN/RB/RO MeSH tree related terms is 1024.8929.

5. Set up the Path Factor for this item based on what paths recommended the item. If MetaMap or Trigram recommended the item AND the item was recommended by PubMed Related Citations, Path Factor equals two, otherwise it equals 1. MetaMap and PubMed Related Citations both recommended our example "Blood Flow Velocity" so the Path Factor is equal to two.

score = score * Path Factor (F)

```
score = 1024.8929 * 2 == 2049.7858
```

The score at the end of processing the Path Factor is 2049.7858.

6. Factor in the item's TermWeight for the final RankScore.

score = score * Item's TermWeight

```
mt_table[5]:  Blood Flow Velocity|C0005798|0|1.6625
score = 2049.7858 * 1.6625 == 3407.7688
```

The final RankScore at the end of processing is 3407.7688. This number than gets truncated (not rounded) to 3407.

The final summarized, clustered, and rank scored (not ordered by score at this point) will look similar to the following:

```
mt_table[0]:   DNA-Binding Proteins|C0012940|1478|1.0150
mt_table[1]:   Transcription Factors|C0040648|1707|1.0150
mt_table[2]:   SEF1 protein|C0212321|207|1.0150
mt_table[3]:   Blood Circulation Time|C0919393|2411|1.1564
mt_table[4]:   Radionuclide Imaging|C0034606|260|1.1564
mt_table[5]:   Blood Flow Velocity|C0005798|3407|1.6625
mt_table[6]:   Neurology|C0027855|89|0.4025
. . .
mt_table[84]:  Confusion|C0009676|55|0.1651
mt_table[85]:  Glasgow Coma Scale|C0017594|246|0.3287
mt_table[86]:  Predictive Value of Tests|C0032944|32|0.1651
mt_table[87]:  Regional Blood Flow|C0034965|118|0.1651
mt_table[88]:  Regression Analysis|C0034980|22|0.1651
```

4. <u>Emphasize Titles</u>

❖ Uses MeSH Terms list

❖ Done for all MeSH Terms in the list

MeSH terms that are identified to be from the Title section of the processed text have their score boosted via the following formula:

    score = current score + (current_score * 2)

5. <u>Emphasize HSTAR (Optional)</u>

❖ Uses MeSH Terms list

❖ Done for all MeSH Terms in the list

❖ Only done when user specifically requests this type of score boosting.

MeSH terms that are identified to be from one of the following MeSH tree hierarchies:

- N01 – N05
- G02 – G03
- L01

have their score boosted via the following formula:

    score = current score + (current_score * HSTAR_FACTOR)

Where HSTAR_FACTOR is the multiplier specified by the user. We are currently using 20 with limited success.

6. Float Chemicals

- ❖ Uses MeSH Terms list
- ❖ Done for all MeSH Terms in the list

Make all chemical (NM) terms score greater than the highest scoring MeSH Heading Mapped to via Restrict to MeSH. If term is NM, then run the term through Restrict to MeSH and receive a list of MeSH Headings that it is Mapped to (HM). We then find the highest scoring HM that is associated with this NM term and set the NM term's score to the highest score plus one.

The following example illustrates how chemicals (NMs) are "floated" up in the MeSH Term list:

Example:

Given the following list after Clustering (list shows MeSH term and associated initial score):

1. Pyrones|26688
2. tipranavir|22812
3. Biological Availability|21954
4. Antacids|20988
5. Pyridines|17301
6. HIV Protease Inhibitors|3077
7. Food|8376
8. Food-Drug Interactions|2379
9. Magnesium Hydroxide|1735
10. Fasting|1598
11. Aluminum Hydroxide|1273
12. Protease Inhibitors|3756
13. Administration, Oral|1048
14. Area Under Curve|1000
15. Indinavir|737
16. Dietary Fats|594
17. Capsules|573
18. Acetamides|510
19. Eating|407
20. Piperidines|381
21. Intestinal Absorption|359
22. Tablets|355
23. Fats|349
24. Tetrazoles|288
25. Pharmacokinetics|273
26. Butyrophenones|272
27. Cross-Over Studies|244
28. Drug Interactions|234
29. Anti-Infective Agents|217
30. Ketoprofen|209
31. Magnesium|193
32. Biphenyl Compounds|184
33. Voluntary Workers|537
34. Chromatography, High Pressure Liquid|162
35. Antiviral Agents|159
36. Analysis of Variance|151
37. Analgesics, Non-Narcotic|149
38. Half-Life|139
39. Histamine H1 Antagonists|128
40. Tromethamine|117
41. Antihypertensive Agents|117
42. Gastric Acidity Determination|116
43. Absorption|112
44. Drug Administration Schedule|109
45. aluminum magnesium hydroxide|102
46. Sedatives, Nonbarbiturate|102

#2 tipranavir with initial score of 22,812:

Restrict to MeSH provides the following list of Headings Mapped to (HM):
- Pyridines with a score of 17,301
- Pyrones with a score of 26,688

Final score for tipranavir becomes 26,689 (score of highest scoring Term (Pyrones|26,688) plus one.

#45 aluminum magnesium hydroxide with initial score of 102:

Restrict to MeSH provides the following list of Headings Mapped to (HM):
- Aluminum Hydroxide with a score of 1,273
- Drug Combinations which is not in the list
- Magnesium Hydroxide with a score of 1,735

Final score for aluminum magnesium hydroxide becomes 1,736 (score of highest scoring Term (Magnesium Hydroxide|1,735) plus one.

7. <u>Determine TopN Terms List</u>

- ❖ Uses freshly sorted MeSH Terms list
- ❖ Done for TopN MeSH Terms in the list only

We want to find the TopN MeSH Heading (MH) terms in our sorted by score list ignoring CheckTags (CT) and SubHeadings (SH) terms in the list. We are going to ignore the CTs and SHs because they are handled separately.

<u>Example:</u>

The user requests TopN to be 25.

1. If we find a CT at positions 7 and 15 in the first 25 terms, we increment TopN for each occurrence (two in this case) to make it 27.

2. If we than find a SH at position 26 in the list, we increment TopN by one which gives us 28.

3. So, the TopN that we will use for the remainder of processing is actually 28 because we want to ignore the CTs at position 7 and 15 and ignore the SH at position 26.

8. <u>Tweak #2 (Optional)</u>

- ❖ Uses MeSH Terms list
- ❖ Done for TopN MeSH Terms in the list only.
- ❖ Only done when user specifically requests this type of term evaluation and processing.

Tweak #2 involves considering the specificity in hierarchies, retaining and removing Terms in the TopN based on their MeSH tree codes. The retaining and removing are done based on several Exceptions (see *Appendix-A*) and Heuristics (see *Appendix-B*) which are processed in the following hierarchical order for the TopN terms in the list:

1. Calculate word counts for each term – to be used in the strcheck Exception.
2. Check Exemptions found in Heuristic #1 a-e, g, and then f.
3. Determine Exceptions 0, A-G.
4. Remove Terms based on Heuristic #2, #3a, #3b, #4, #5, #6, #7, #8, #9, #10 in order of Heuristic number.

9. <u>MH/SH Substitution</u>

- ❖ Uses MeSH Terms list
- ❖ Done for TopN MeSH Terms in the list, which survived the Tweak #2 removal process only.

If a Term is a MeSH Heading (MH) and there is a corresponding SubHeading (SH), only show the SubHeading Term.

1. First we look for a direct match in the TopN MH Terms of a SubHeading anywhere else in the list. If we find a match, we are done looking and the substitution takes place. e.g., MH of "Pharmacokinetics" becomes SH of "pharmacokinetics".

2. If we don't find a direct match above, we go through the supplemental MH/SH Lookup list (***Appendix-F***). If we find a match from the lookup list, the substitution takes place.

10. <u>Validate TopN Terms</u>

- ❖ Uses MeSH Terms list
- ❖ Done for TopN MeSH Terms in the list, which survived the Tweak #2 removal process and the MH/SH Substitution process only.

For each TopN MeSH Term which is a MH or CT the following tests are ran in order with their corresponding additions of CheckTags and SubHeadings taking place when appropriate.

if CUI = C0042542 (Vero Cells), then add MH+ of Cercopithecus aethiops.

if CUI = C0085080 (Chinese hamster ovary cell), then add CT of Hamster AND Animal.

if CUI is in Adolescence list, then add CT of Adolescence AND CT of Human.

if CUI is in Aged list, then add CT of Aged AND CT of Human.

if CUI is in Animal list, then add CT of Animal.

if CUI is in Cattle list, then add CT of Cattle AND CT of Animal.

if CUI is in Cat list, then add CT of Cat AND CT of Animal.

if CUI is in Dog list, then add CT of Dog AND CT of Animal.

if CUI is in Female list, then add CT of Female.

if CUI is in Human list, then add CT of Human.

if CUI is in Newborn list, then add CT of Infant, Newborn AND CT of Human.

if CUI is in Male list, then add CT of Male.

if CUI is in Pregnant list, then add CT of Pregnancy AND if Female hasn't already been added, add Female.

if CUI is in Mice list and recommendation is from MetaMap, then add CT of Mice AND Animal.

if CUI is in Rats list and recommendation is from MetaMap, then add CT of Rats AND Animal.

if CUI is in Sheep list and recommendation is from MetaMap, then add MH+ of Sheep AND CT of Animal.

if CUI is in Swine list and recommendation is from MetaMap, then add MH+ of Swine AND CT of Animal.

if CUI is in United States list and recommendation is from MetaMap, then add MH+ of United States.

if Male CT not used AND this concept has a tree code found in the Male tree list, add CT of Male.

if Mice CT not used AND recommendation is from MetaMap, AND this concept has a tree code found in the Mice tree list, add CT of Mice AND add CT of Animal.

if Rat CT not used AND recommendation is from MetaMap, AND this concept has a tree code found in the Rat tree list, add CT of Rat AND add CT of Animal.

if Female CT not used AND this concept has a tree code found in the Female tree list, add CT of Female.

if Pregnancy CT not used AND this concept has a tree code found in the Pregnancy tree list, add CT of Pregnancy AND add CT of Female.

if Infant, Newborn CT not used AND this concept has a tree code found in the Newborn tree list, add CT of Infant, Newborn AND add CT of Human.

if Animal CT not used AND this concept has a tree code found in the Animal tree list, add CT of Animal.

if Aged CT not used AND this concept has a tree code found in the Aged tree list, add CT of Aged AND add CT of Human.

if Human CT not used AND this concept has a tree code found in the Human tree list, add CT of Human.

if Hamster CT not used AND recommendation is from MetaMap, AND this concept has a tree code found in the Hamster tree list, add CT of Hamster AND add CT of Animal.

if United States MH+ not used AND this concept has a tree code found in the United States tree list, add MH+ of United States.

if the concept's tree code is in "G05", add SH "genetics"

else if the concept's tree code is in "G04.610", add SH "immunology"

else if the concept's tree code is in"G03.850.310", add SH "transmission"

else if the concept's tree code is in"G12.091.690.140", add SH "pharmacokinetics"

else if the concept's tree code is in"G04.185.515.880", add SH "virology"

else if the concept's tree code is in"E01.370.350.700", add SH "radiography"

else if the concept's tree code is in"E01.370.384.730", add SH "radionuclide imaging"

else if the concept's tree code is in"E01.370.350.850", add SH "ultrasonography"

else if the concept's tree code is in"E02.810" AND NOT in "E02.810.530", add SH "radiotherapy"

else if the concept's tree code is in"E02.831", add SH "rehabilitation"

else if the concept's tree code is in"E04.936", add SH "transplantation"

else if the concept's tree code is in"E04 AND NOT in "E01", add SH "surgery"

else if the concept's tree code is in"N03.219", add SH "economics"

11. **Add "drug therapy" SH**

- ❖ Uses MeSH Terms list
- ❖ Done for TopN MeSH Terms in the list, which survived the Tweak #2 removal process and the MH/SH Substitution process only.

When therapy (SH) has been recommended AND either: the concept's tree code is in E02.319, except E02.319.703 OR A term has been exempted on account of Heuristic #1f - add SH of "drug therapy".

12. **Drop "physiology" & "analysis" SHs**

- ❖ Uses MeSH Terms list
- ❖ Done for TopN MeSH Terms in the list, which survived the Tweak #2 removal process and the MH/SH Substitution process only.

- Remove physiology (SH) unless some term in topN is in Categories G04-G11.

- Remove analysis (SH) unless some term in topN is in Categories D01-D25 OR in E05.196 OR in H01.181.278.

13. <u>Add CheckTags from Text</u>

❖ Uses the Title and Abstract fields from the actual text of the citation.

❖ Done for all CheckTag substitutions found in the CheckTag Lookup list (***Appendix-D***).

If a CheckTag in the lookup list is found in the text of either the Title or the Abstract, we verify that the CheckTag has not already been added as a result and if it hasn't, we add it. Care is taken to make sure that CheckTags in the lookup table map to actual words in the text and are not part of other words.

14. <u>Add Geographics from Text</u>

❖ Uses the Title and Abstract fields from the actual text of the citation.

❖ Done for all Geographic substitutions found in the Geographics Lookup list (***Appendix-E***).

If a Geographic in the lookup list is found in the text of either the Title or the Abstract, we verify that the Geographic has not already been added as a result and if it hasn't, we add it. Care is taken to make sure that Geographic in the lookup table map to actual words in the text and are not part of other words.

15. <u>Display Results</u>

❖ Uses MeSH Terms list

❖ Done for TopN MeSH Terms in the list that have survived the Tweak #2 removal process and the MH/SH Substitution process only.

In the display results section of the program, we still have the potential of adding new terms based on items found in the entire MeSH Term list (see step 3 below). For all but Step #3 below, we only use the TopN MeSH Terms. The list below details the ordering of how we print out the final results of the MTI.

1. For each of the TopN Terms in the list that are oktoprint (not removed due to tweaks or substitution) and are either MH or NM – we print out the result.

2. Print "------"

3. For Terms below TopN (TopN + 1 – end of list), we check to see if they might be any "Special Terms". Where "Special Terms" are terms deemed "special" by the program and include MHs that are out of the scope of our normal recommendation scoring by virtue of being scored lower than TopN. They must have a tree code that falls within one of the following trees: "Z01" [except Z01.433 (Cities) and Z01.586 (Historical Geographic Locations)], "E05.318.760.500" (Epidemiologic Studies), and "G03.850.520.835.500" (Epidemiologic Studies).
These terms are printed using a "MH-S" denotation.

4. Print out Other Terms – which are derived through the validation rules. These terms are printed using a "MH+" denotation.

5. Print out CheckTags found in the TopN.

6. Print out CheckTags added through validation, etc. not duplicated in TopN entries.

7. Print out SubHeadings found in TopN AND have a score > 200.

8. Print out SubHeadings added through validation that are not duplicates of TopN entries

Appendix A – MTI Exceptions for Tweak #2

strcheck   A more general term can be a term of which the term to be acted upon is a sublist. (e.g., "Life" is more general than "Quality of Life" and "Quality of Life" is more specific than "Life"). Strcheck is not used when the difference in number of words is greater than 4, or on Heuristic #7.

0. When Surgery (MH) is in topN treat it as being E04 as well. (Add E04 to the tree code list for Surgery)

A. When a term is in E03 (Anesthesia and Analgesia), consider the term to be more specific than the top-level term in E02 (Therapeutics).

B. When a term is in E04 (Surgical Procedures, Operative) and not also in E01 (Diagnosis), consider the term to be more specific than the top-level term in E02 (Therapeutics).

C. When a term is in both E01 (Diagnosis) and E04 (Surgical Procedures, Operative), ignore any and all E04 tree numbers.

D. When a term is in both E01 (Diagnosis) and E05 (Investigative Techniques), ignore any and all E05 tree numbers.

E. When a term is in G01 (Biological Sciences) or G02 (Health Occupations), and also in some other subcategory, ignore any and all G01 and G02 tree numbers.

F. When a term is in E04 (Surgical Procedures, Operative), consider the term to be more specific than G02.403.810.762.

G. When a term has tree number G02.403.810.762 consider the term to be more specific than the top-level term in E02 (Therapeutics).

Appendix B – MTI Heuristics for Tweak #2

- Heuristic #1 items to be exempted:
    a. If the same term has been assigned by both methods, keep it.
    b. Don't remove NMs
    c. Don't remove SHs
    d. Keep items in Geographics (Z01) tree recommended by MetaMap
    e. Keep MH items in topN that are substitutes for SHs when recommended by MetaMap only
    f. When a term meets the following criteria, mark it as exempt:
        1. If we have any item in topN that is in range of D01-D25 AND NOT in range D26-D27 AND Recommended by both MM AND RC proceed,
        2. For each term in the topN AND in the range of D01-D25 AND in the range of D26-D27 AND not in any other categories outside of these Dnn categories AND is recommended by the RC path only, proceed for each term fitting this criteria,
        3. For each term, compile a list of all descendant terms in the topN that are in the range of D01-D27 AND NOT IN ANY other category including this term and find the highest scoring item in this pool of terms and mark it as exempt.
    g. Keep CT items recommended by MetaMap only.

- Heuristic #2: For removing terms when the method is MetaMap only. Remove terms resulting from Restrict to MeSH having no Semantic Type (ST) in common with the set of Semantic Types for the concepts that were recommended by MetaMap before Restrict to MeSH was run.

- Heuristic #3 is for removing terms when the method is Related Citations only. Remove the following:
    a. Check Tags
    b. Geographics (Z01)

- Heuristic #4 is for removing terms when the method is Related Citations only. Remove any term in range D01 - D25 AND in range D26 - D27 AND NOT in any other categories AND when MetaMap hasn't recommended any term within range D01 - D25.

- Heuristic #5 is for removing terms when the method is MetaMap only. In any instance, when MM assigns a term which is more general or more specific OR RC assigns a term which is more specific, add the term to a collection of kept terms. When the collection is not empty, remove a term when it is not a member of the collection of kept terms AND the score for the term is lower than the lowest-scoring term in the collection of kept terms. The removal is not done unless the lowest_scoring term in the collection is less than 10,000.

- Heuristic #6 is for removing terms when only MetaMap recommends a term. In any instance, when both methods assign a more specific term, remove the term. **Note:** Uses Exceptions from above.

- Heuristic #7 is for removing terms when the method is Related Citations only. In every instance, when MetaMap assigns no term which is more general, remove the term. **Note:** Uses Exceptions from above.

- Heuristic #8 is for removing terms when the method is Related Citations only. In any instance, when both methods assign a more general term, remove the term. **Note:** Uses Exceptions from above.

- Heuristic #9 is for removing terms when only MetaMap recommends a term. In any instance, when there is no RC term from the same category, remove the term - provided the term is not in categories H or I.

- Heuristic #10 is for removing terms when the method is MetaMap only. Remove any term which is an NM when there are no terms that are a Heading Mapped-to (HM).

Appendix C – Lookup Lists

```
Adolescence (6):
        1) C0001580 - Adolescent Behavior
        2) C0001585 - Adolescent, Hospitalized
        3) C0001586 - Adolescent, Institutionalized
        4) C0032968 - Pregnancy in Adolescence
        5) C0085100 - Adolescent Health Services
        6) C0162630 - Adolescent Nutrition

------------------------------------

Aged (5):
        1) C0001795 - Aged, 80 and over
        2) C0013772 - Elder Abuse
        3) C0018753 - Health Services for the Aged
        4) C0019870 - Homes for the Aged
        5) C0079204 - Dental Care for Aged

Aged Trees (1):
        M01.060.116.100

------------------------------------

Animal (206):
        1) C0000780 - Abomasum
        2) C0000823 - Abortion, Veterinary
        3) C0001247 - Actinobacillosis
        4) C0001748 - African Horse Sickness
        5) C0001752 - African Swine Fever
        6) C0001878 - Air Sacs
        7) C0002016 - Aleutian Mink Disease
        8) C0002757 - Anal Gland Neoplasms
        9) C0002759 - Anal Sacs
       10) C0002797 - Anaplasmosis
       11) C0003046 - Animal Communication
       12) C0003047 - Animal Diseases
       13) C0003054 - Animal Nutrition
       14) C0003452 - Antlers
       15) C0004421 - Avian Leukosis
       16) C0004426 - Sarcoma, Avian
       17) C0004576 - Babesiosis
       18) C0004895 - Beak
       19) C0004935 - Behavior, Animal
       20) C0005591 - Bird Diseases
       21) C0005866 - Bluetongue
       22) C0006008 - Border Disease
       23) C0006023 - Borna Disease
       24) C0006075 - Bovine Virus Diarrhea-Mucosal Disease
       25) C0006311 - Brucellosis, Bovine
       26) C0006440 - Bursa of Fabricius
       27) C0007098 - Carcinoma 256, Walker
       28) C0007122 - Carcinoma, Brown-Pearce
       29) C0007125 - Carcinoma, Ehrlich Tumor
       30) C0007128 - Carcinoma, Krebs 2
       31) C0007288 - Carpus, Animal
       32) C0007350 - Cat Diseases
       33) C0007450 - Cats
       34) C0007452 - Cattle
       35) C0007453 - Cattle Diseases
       36) C0008046 - Chick Embryo
       37) C0009424 - Comb and Wattles
       38) C0009990 - Copulation
       39) C0010085 - Corpora Allata
       40) C0010352 - Crop, Avian
       41) C0010418 - Cryptosporidiosis
       42) C0011853 - Diabetes Mellitus, Experimental
```

```
Animal (continued):
         43) C0012118 - Dictyocaulus Infections
         44) C0012602 - Dirofilariasis
         45) C0012754 - Distemper
         46) C0012979 - Dog Diseases
         47) C0012984 - Dogs
         48) C0013076 - Dourine
         49) C0013529 - Echolocation
         50) C0013570 - Ecthyma, Contagious
         51) C0013591 - Ectromelia, Infectious
         52) C0013605 - Edema Disease of Swine
         53) C0013702 - Egg Shell
         54) C0013782 - Electric Organ
         55) C0013897 - Eliminative Behavior, Animal
         56) C0013940 - Embryo, Nonmammalian
         57) C0014073 - Encephalomyelitis, Enzootic Porcine
         58) C0014342 - Enteritis, Transmissible, of Turkeys
         59) C0014371 - Enterotoxemia
         60) C0014481 - Ephemeral Fever
         61) C0014521 - Epidermitis, Exudative, of Swine
         62) C0014661 - Equine Infectious Anemia
         63) C0014736 - Erysipelothrix Infections
         64) C0015655 - Fascioloidiasis
         65) C0015665 - Fat Body
         66) C0015731 - Feathers
         67) C0015765 - Feline Panleukopenia
         68) C0016154 - Fish Diseases
         69) C0016513 - Foot Rot
         70) C0016514 - Foot-and-Mouth Disease
         71) C0016555 - Forelimb
         72) C0016627 - Fowl Plague
         73) C0016629 - Fowlpox
         74) C0016697 - Freemartinism
         75) C0017162 - Gastroenteritis, Transmissible, of Swine
         76) C0017558 - Gills
         77) C0017584 - Gizzard
         78) C0017589 - Glanders
         79) C0018018 - Goat Diseases
         80) C0018249 - Grooming
         81) C0018382 - Guinea Pigs
         82) C0018557 - Hamsters
         83) C0018597 - Harderian Gland
         84) C0018835 - Heartwater Disease
         85) C0018891 - Helminthiasis, Animal
         86) C0019051 - Hemolymph
         87) C0019188 - Hepatitis, Animal
         88) C0019191 - Hepatitis, Infectious Canine
         89) C0019194 - Hepatitis, Viral, Animal
         90) C0019549 - Hindlimb
         91) C0019556 - Hip Dysplasia, Canine
         92) C0019841 - Hog Cholera
         93) C0019861 - Homing Behavior
         94) C0019909 - Hoof and Claw
         95) C0019939 - Horns
         96) C0019940 - Horse Diseases
         97) C0021334 - Infectious Bovine Rhinotracheitis
         98) C0021800 - Interrenal Gland
         99) C0022576 - Keratoconjunctivitis, Infectious
        100) C0022976 - Lameness, Animal
        101) C0023420 - Leukemia L1210
        102) C0023421 - Leukemia L5178
        103) C0023429 - Leukemia P388
        104) C0023904 - Liver Neoplasms, Experimental
        105) C0024003 - Lordosis
        106) C0024025 - Louping Ill
        107) C0024106 - Lumpy Skin Disease
        108) C0024533 - Malaria, Avian
        109) C0024587 - Malignant Catarrh
```

```
Animal (continued):
        110) C0024648 - Malpighian Tubules
        111) C0024659 - Mammae
        112) C0024667 - Mammary Neoplasms
        113) C0024668 - Mammary Neoplasms, Experimental
        114) C0024788 - Marburg Virus Disease
        115) C0024793 - Marek's Disease
        116) C0024895 - Mastitis, Bovine
        117) C0025864 - Metrial Gland
        118) C0025914 - Mice
        119) C0026131 - Milk
        120) C0026414 - Monieziasis
        121) C0026431 - Monkey Diseases
        122) C0026851 - Muscular Dystrophy, Animal
        123) C0027152 - Myxomatosis, Infectious
        124) C0027345 - Nairobi Sheep Disease
        125) C0027776 - Nesting Behavior
        126) C0027983 - Newcastle Disease
        127) C0028058 - Nictitating Membrane
        128) C0028972 - Omasum
        129) C0029129 - Optic Lobe
        130) C0029954 - Oviducts
        131) C0030209 - Pair Bond
        132) C0030500 - Parasitic Diseases, Animal
        133) C0030524 - Paratuberculosis
        134) C0030612 - Parturient Paresis
        135) C0031021 - Perianal Glands
        136) C0032243 - Pleuropneumonia, Contagious
        137) C0032291 - Pneumonia, Atypical Interstitial, of Cattle
        138) C0032306 - Pneumonia, Progressive Interstitial, of Sheep
        139) C0032851 - Poultry Diseases
        140) C0032942 - Predatory Behavior
        141) C0033741 - Protozoan Infections, Animal
        142) C0033745 - Proventriculus
        143) C0033839 - Pseudorabies
        144) C0034049 - Pulmonary Adenomatosis, Ovine
        145) C0034493 - Rabbits
        146) C0034531 - Radiation Injuries, Experimental
        147) C0034693 - Rats
        148) C0035295 - Reticulum
        149) C0035613 - Rift Valley Fever
        150) C0035637 - Rinderpest
        151) C0035801 - Rodent Diseases
        152) C0035946 - Rumen
        153) C0036118 - Salmonella Infections, Animal
        154) C0036139 - Salt Gland
        155) C0036294 - Scent Glands
        156) C0036457 - Scrapie
        157) C0036850 - Setariasis
        158) C0036865 - Sex Behavior, Animal
        159) C0036946 - Sheep Diseases
        160) C0036969 - Pasteurellosis, Pneumonic
        161) C0038235 - Steatitis
        162) C0038328 - Stifle
        163) C0038360 - Stomach, Avian
        164) C0038361 - Stomach, Ruminant
        165) C0038459 - Strongyle Infections, Equine
        166) C0038981 - Swayback
        167) C0039006 - Swine Diseases
        168) C0039007 - Swine Erysipelas
        169) C0039010 - Swine Vesicular Disease
        170) C0039259 - Tail
        171) C0039325 - Tarsus, Animal
        172) C0039753 - Theileriasis
        173) C0040553 - Toxocariasis
        174) C0040559 - Toxoplasmosis, Animal
        175) C0041230 - Trypanosomiasis, Bovine
        176) C0041306 - Tuberculosis, Avian
```

```
Animal (continued):
        177) C0041307 - Tuberculosis, Bovine
        178) C0041605 - Ultimobranchial Body
        179) C0042465 - Venereal Tumors, Veterinary
        180) C0042542 - Vero Cells
        181) C0042567 - Vertebrates
        182) C0042584 - Vesicular Exanthema of Swine
        183) C0042640 - Vibrissae
        184) C0042932 - Vocalization, Animal
        185) C0043153 - White Muscle Disease
        186) C0043189 - Wing
        187) C0043220 - Wool
        188) C0043528 - Zoonoses
        189) C0079335 - Feline Acquired Immunodeficiency Syndrome
        190) C0079864 - Murine Acquired Immunodeficiency Syndrome
        191) C0080151 - Simian Acquired Immunodeficiency Syndrome
        192) C0080323 - Visna
        193) C0085164 - Leukemia, Feline
        194) C0085165 - Enzootic Bovine Leukosis
        195) C0085209 - Encephalopathy, Bovine Spongiform
        196) C0085262 - PC12 Cells
        197) C0085306 - Feline Infectious Peritonitis
        198) C0206436 - Photoreceptors, Invertebrate
        199) C0242598 - LLC-PK1 Cells
        200) C0242634 - Primate Diseases
        201) C0242635 - Ape Diseases
        202) C0243038 - Carcinoma, Lewis Lung
        203) C0376538 - Porcine Reproductive and Respiratory Syndrome
        204) C0376702 - COS Cells
        205) C0518461 - Grooming self-care
        206) C0600243 - Home Range

-----------------------------------

Cattle (15):
         1) C0006075 - Bovine Virus Diarrhea-Mucosal Disease
         2) C0006311 - Brucellosis, Bovine
         3) C0007453 - Cattle Diseases
         4) C0014481 - Ephemeral Fever
         5) C0016697 - Freemartinism
         6) C0021334 - Infectious Bovine Rhinotracheitis
         7) C0024106 - Lumpy Skin Disease
         8) C0024587 - Malignant Catarrh
         9) C0024895 - Mastitis, Bovine
        10) C0032291 - Pneumonia, Atypical Interstitial, of Cattle
        11) C0039753 - Theileriasis
        12) C0041230 - Trypanosomiasis, Bovine
        13) C0041307 - Tuberculosis, Bovine
        14) C0085165 - Enzootic Bovine Leukosis
        15) C0085209 - Encephalopathy, Bovine Spongiform

-----------------------------------

Cat (4):
         1) C0007350 - Cat Diseases
         2) C0079335 - Feline Acquired Immunodeficiency Syndrome
         3) C0085164 - Leukemia, Feline
         4) C0085306 - Feline Infectious Peritonitis

-----------------------------------

Dog (3):
         1) C0012979 - Dog Diseases
         2) C0019191 - Hepatitis, Infectious Canine
         3) C0019556 - Hip Dysplasia, Canine
```

```
Female (171):
         1) C0000806 - Abortion, Eugenic
         2) C0000811 - Abortion, Induced
         3) C0000812 - Abortion, Legal
         4) C0000820 - Abortion, Therapeutic
         5) C0001575 - Adnexa Uteri
         6) C0001576 - Adnexal Diseases
         7) C0001577 - Adnexitis
         8) C0002453 - Amenorrhea
         9) C0002935 - Anestrus
        10) C0003128 - Anovulation
        11) C0004768 - Bartholin's Glands
        12) C0006205 - Broad Ligament
        13) C0006852 - Candidiasis of vagina
        14) C0007860 - Cervicitis
        15) C0007867 - Cervix Diseases
        16) C0007868 - Cervix Dysplasia
        17) C0007869 - Cervix Erosion
        18) C0007871 - Cervix Incompetence
        19) C0007873 - Cervix Neoplasms
        20) C0007874 - Cervix Uteri
        21) C0007876 - Cesarean Section
        22) C0008043 - Chiari-Frommel Syndrome
        23) C0008984 - Clitoris
        24) C0010092 - Corpus Luteum
        25) C0010096 - Corpus Luteum Regression
        26) C0011106 - Decidua
        27) C0011209 - Delivery
        28) C0012154 - Diestrus
        29) C0012358 - Dilatation and Curettage
        30) C0013390 - Dysmenorrhea
        31) C0013394 - Dyspareunia
        32) C0014170 - Endometrial Neoplasms
        33) C0014173 - Endometrial Hyperplasia
        34) C0014175 - Endometriosis
        35) C0014179 - Endometritis
        36) C0014180 - Endometrium
        37) C0014586 - Episiotomy
        38) C0014935 - Estrogen Replacement Therapy
        39) C0014948 - Estrus
        40) C0014950 - Estrus Synchronization
        41) C0015362 - Extraction, Obstetrical
        42) C0015556 - Fallopian Tube Diseases
        43) C0015558 - Fallopian Tube Neoplasms
        44) C0015560 - Fallopian Tubes
        45) C0016426 - Follicular Atresia
        46) C0016431 - Follicular Fluid
        47) C0016434 - Follicular Phase
        48) C0016722 - Frigidity
        49) C0016941 - Galactorrhea
        50) C0016999 - Gamete Intrafallopian Transfer
        51) C0017411 - Genital Diseases, Female
        52) C0017416 - Genital Neoplasms, Female
        53) C0017421 - Genitalia, Female
        54) C0018120 - Ovarian Follicle
        55) C0018207 - Granulosa Cells
        56) C0018414 - Gynatresia
        57) C0018934 - Hematocolpos
        58) C0018948 - Hematometra
        59) C0019857 - Home Childbirth
        60) C0020412 - Hymen
        61) C0020699 - Hysterectomy
        62) C0020700 - Hysterectomy, Vaginal
        63) C0021361 - Infertility, Female
        64) C0022783 - Kraurosis Vulvae
        65) C0022875 - Labor, Induced
        66) C0022925 - Lactation
        67) C0022927 - Lactation Disorders
```

```
Female (continued):
         68) C0023372 - Homosexuality, Female
         69) C0023533 - Leukorrhea
         70) C0024153 - Luteal Phase
         71) C0024156 - Lutein Cells
         72) C0024894 - Mastitis
         73) C0024895 - Mastitis, Bovine
         74) C0025184 - Meigs' Syndrome
         75) C0025274 - Menarche
         76) C0025320 - Menopause
         77) C0025322 - Menopause, Premature
         78) C0025323 - Menorrhagia
         79) C0025329 - Menstrual Cycle
         80) C0025344 - Menstruation
         81) C0025345 - Menstruation Disturbances
         82) C0025597 - Metestrus
         83) C0025874 - Metrorrhagia
         84) C0026132 - Milk Ejection
         85) C0027088 - Myometrium
         86) C0027484 - Natural Childbirth
         87) C0028949 - Oligomenorrhea
         88) C0029051 - Oophoritis
         89) C0029458 - Osteoporosis, Postmenopausal
         90) C0029927 - Ovarian Cysts
         91) C0029928 - Ovarian Diseases
         92) C0029936 - Ovariectomy
         93) C0029939 - Ovary
         94) C0029957 - Oviposition
         95) C0029965 - Ovulation
         96) C0029976 - Ovum Implantation
         97) C0029977 - Ovum Implantation, Delayed
         98) C0029979 - Ovum Transport
         99) C0030455 - Parametritis
        100) C0030563 - Parity
        101) C0030584 - Parovarian Cyst
        102) C0032460 - Polycystic Ovary Syndrome
        103) C0032797 - Postpartum Hemorrhage
        104) C0032961 - Pregnancy
        105) C0032986 - Pregnancy, Animal
        106) C0033046 - Premenstrual Syndrome
        107) C0033274 - Proestrus
        108) C0033778 - Pruritus Vulvae
        109) C0033831 - Pseudopregnancy
        110) C0034040 - Puerperal Disorders
        111) C0034041 - Puerperal Infection
        112) C0034042 - Puerperium
        113) C0034895 - Rectovaginal Fistula
        114) C0035877 - Round Ligament
        115) C0036130 - Salpingitis
        116) C0036136 - Salpingostomy
        117) C0037853 - Sperm-Ovum Interactions
        118) C0038289 - Sterilization, Tubal
        119) C0038835 - Superovulation
        120) C0038902 - Gynecologic Surgical Procedures
        121) C0038906 - Obstetric Surgical Procedures
        122) C0039748 - Theca Cells
        123) C0040923 - Trichomonas Vaginitis
        124) C0041311 - Tuberculosis, Female Genital
        125) C0042130 - Uterine Contraction
        126) C0042131 - Uterine Diseases
        127) C0042134 - Uterine Hemorrhage
        128) C0042138 - Uterine Neoplasms
        129) C0042139 - Uterine Perforation
        130) C0042140 - Uterine Prolapse
        131) C0042143 - Uterine Rupture
        132) C0042149 - Uterus
        133) C0042223 - Vacuum Curettage
        134) C0042225 - Vacuum Extraction, Obstetrical
```

```
Female (continued):
        135) C0042232 - Vagina
        136) C0042251 - Vaginal Diseases
        137) C0042253 - Vaginal Fistula
        138) C0042258 - Vaginal Neoplasms
        139) C0042261 - Vaginal Smears
        140) C0042267 - Vaginitis
        141) C0042556 - Version, Fetal
        142) C0042582 - Vesicovaginal Fistula
        143) C0042993 - Vulva
        144) C0042994 - Vulvar Diseases
        145) C0042995 - Vulvar Neoplasms
        146) C0042996 - Vulvitis
        147) C0042998 - Vulvovaginitis
        148) C0043209 - Woman
        149) C0043210 - Women
        150) C0079341 - Circumcision, Female
        151) C0080301 - Vaginal Birth after Cesarean
        152) C0080339 - Women's Health
        153) C0085076 - Mammaplasty
        154) C0085083 - Ovarian Hyperstimulation Syndrome
        155) C0085166 - Vaginosis, Bacterial
        156) C0085215 - Ovarian Failure, Premature
        157) C0162482 - Uterine Inversion
        158) C0206076 - Reproductive History
        159) C0206101 - Cesarean Section, Repeat
        160) C0206158 - Premenopause
        161) C0206159 - Postmenopause
        162) C0221074 - Depression, Postpartum
        163) C0227791 - Vaginal Discharge
        164) C0242810 - Battered Women
        165) C0242836 - Pregnancy Reduction, Multifetal
        166) C0243033 - Maternal Exposure
        167) C0269886 - Inversion of uterus during delivery
        168) C0269995 - Galactorrhea associated with childbirth
        169) C0392535 - Diagnosis of induced abortion
        170) C0496920 - Ovarian Neoplasms
        171) C0600454 - Cervical Ripening

Female Trees (12):
        A05.360.319
        C13.371.56
        C23.550.568
        C13.371.852
        C13.371.894
        C13.371.944
        C13.703.844.253
        G08.520.440
        G08.520.780.218
        G08.520.882
        E04.520
        E04.950.300
```

```
Human (27):
        1) C0001578 - Adolescence
        2) C0001587 - Adolescent
        3) C0001675 - Adult
        4) C0001792 - Aged
        5) C0008059 - Child
        6) C0008100 - Child, Preschool
        7) C0016539 - Forefoot, Human
        8) C0017429 - Genome, Human
        9) C0018873 - Hela Cells
       10) C0019874 - Hominidae
       11) C0021270 - Infant
       12) C0021289 - Infant, Newborn
       13) C0022539 - KB Cells
       14) C0025266 - Men
       15) C0026062 - Middle Age
       16) C0029458 - Osteoporosis, Postmenopausal
       17) C0030705 - Patients
       18) C0043209 - Woman
       19) C0043210 - Women
       20) C0079204 - Dental Care for Aged
       21) C0079377 - Frail Elderly
       22) C0080339 - Women's Health
       23) C0085429 - Koro
       24) C0282549 - HL-60 Cells
       25) C0282560 - Caco-2 Cells
       26) C0282639 - HT29 Cells
       27) C0376448 - Jurkat Cells

Human Trees (4):
        A11.223.475.480
        M01.060
        A01.378.592.350.300
        A01.378.592.350.510.800

-------------------------------------

Newborn (22):
        1) C0002636 - Amniotic Band Syndrome
        2) C0002891 - Anemia, Neonatal
        3) C0004045 - Asphyxia Neonatorum
        4) C0006287 - Bronchopulmonary Dysplasia
        5) C0014761 - Erythroblastosis, Fetal
        6) C0019088 - Hemorrhagic Disease of Newborn
        7) C0020192 - Hyaline Membrane Disease
        8) C0021290 - Infant, Newborn, Diseases
        9) C0021295 - Infant, Premature, Diseases
       10) C0021709 - Intensive Care Units, Neonatal
       11) C0021711 - Intensive Care, Neonatal
       12) C0022353 - Jaundice, Neonatal
       13) C0023529 - Leukomalacia, Periventricular
       14) C0025048 - Meconium Aspiration
       15) C0027609 - Neonatal Abstinence Syndrome
       16) C0027617 - Neonatal Screening
       17) C0029076 - Ophthalmia Neonatorum
       18) C0031190 - Persistent Fetal Circulation Syndrome
       19) C0035220 - Respiratory Distress Syndrome
       20) C0035344 - Retinopathy of Prematurity
       21) C0036415 - Sclerema Neonatorum
       22) C0079893 - Neonatal Nursing

Newborn Trees (2):
        C16.614.521
        M01.060.392.520.520
```

```
Male (84):
      1) C0001216 - Acrosome
      2) C0004690 - Balanitis
      3) C0006366 - Bulbourethral Glands
      4) C0008819 - Circumcision
      5) C0010417 - Cryptorchidism
      6) C0013746 - Ejaculation
      7) C0013747 - Ejaculatory Ducts
      8) C0014533 - Epididymis
      9) C0014534 - Epididymitis
     10) C0017412 - Genital Diseases, Male
     11) C0017417 - Genital Neoplasms, Male
     12) C0017422 - Genitalia, Male
     13) C0018418 - Gynecomastia
     14) C0018931 - Hematocele
     15) C0020252 - Hydrocele
     16) C0020646 - Hypospadias
     17) C0021116 - Impotence
     18) C0021364 - Infertility, Male
     19) C0023602 - Leydig Cells
     20) C0025266 - Men
     21) C0028960 - Oligospermia
     22) C0029191 - Orchitis
     23) C0030483 - Paraphimosis
     24) C0030846 - Penile Diseases
     25) C0030847 - Penile Erection
     26) C0030848 - Penile Induration
     27) C0030849 - Penile Neoplasms
     28) C0030851 - Penis
     29) C0031538 - Phimosis
     30) C0033117 - Priapism
     31) C0033572 - Prostate
     32) C0033573 - Prostatectomy
     33) C0033575 - Prostatic Diseases
     34) C0033577 - Prostatic Hyperplasia
     35) C0033578 - Prostatic Neoplasms
     36) C0033581 - Prostatitis
     37) C0034919 - Redundant prepuce and phimosis
     38) C0035278 - Rete Testis
     39) C0036471 - Scrotum
     40) C0036628 - Seminal Vesicles
     41) C0036629 - Seminiferous Epithelium
     42) C0036630 - Seminiferous Tubules
     43) C0036770 - Sertoli Cells
     44) C0037839 - Sperm Agglutination
     45) C0037840 - Sperm Banks
     46) C0037841 - Sperm Capacitation
     47) C0037842 - Sperm Count
     48) C0037844 - Sperm Head
     49) C0037846 - Sperm Maturation
     50) C0037848 - Sperm Motility
     51) C0037851 - Sperm Tail
     52) C0037852 - Sperm Transport
     53) C0037853 - Sperm-Ovum Interactions
     54) C0037855 - Spermatic Cord
     55) C0037856 - Spermatic Cord Torsion
     56) C0037857 - Spermatids
     57) C0037859 - Spermatocele
     58) C0037863 - Spermatocytes
     59) C0037864 - Spermatogenesis
     60) C0037866 - Spermatogonia
     61) C0037868 - Spermatozoa
     62) C0038914 - Urologic Surgical Procedures, Male
     63) C0039584 - Testicular Diseases
     64) C0039585 - Testicular Feminization
     65) C0039590 - Testicular Neoplasms
     66) C0039597 - Testis
     67) C0041317 - Tuberculosis, Male Genital
```

```
Male (continued):
        68) C0042341 - Varicocele
        69) C0042360 - Vas Deferens
        70) C0042421 - Vasovasostomy
        71) C0085429 - Koro
        72) C0153604 - Malignant neoplasm of scrotum
        73) C0242658 - Homosexuality, Male
        74) C0242788 - Breast Neoplasms, Male
        75) C0243000 - Impotence, Vasculogenic
        76) C0243034 - Paternal Exposure
        77) C0268896 - Disease of seminal vesicle
        78) C0268919 - Disease of scrotum
        79) C0345326 - Congenital phimosis
        80) C0403766 - Acquired phimosis
        81) C0428003 - Sperm motility measurement
        82) C0600338 - Sperm Number
        83) C0678217 - Encounter due to sperm count
        84) C0700113 - Epididymis disorders


Male Trees (8):
        A05.360.444.849
        C12.294.365.700
        C12.294.494
        C12.294.565
        C12.294.829
        G08.520.310.760
        A11.497.760
        E04.950.774.860

------------------------------------

Pregnant (88):
         1) C0000786 - Abortion, Spontaneous
         2) C0000809 - Abortion, Habitual
         3) C0000810 - Abortion, Incomplete
         4) C0000814 - Abortion, Missed
         5) C0000817 - Abortion, Septic
         6) C0000821 - Abortion, Threatened
         7) C0000823 - Abortion, Veterinary
         8) C0000832 - Abruptio Placentae
         9) C0006157 - Breech Presentation
        10) C0007871 - Cervix Incompetence
        11) C0008493 - Hydatidiform Mole, Invasive
        12) C0008495 - Chorioamnionitis
        13) C0008497 - Choriocarcinoma
        14) C0008509 - Chorionic Villi Sampling
        15) C0010095 - Corpus Luteum Maintenance
        16) C0013418 - Dystocia
        17) C0013537 - Eclampsia
        18) C0013927 - Embolism, Amniotic Fluid
        19) C0015944 - Fetal Membranes, Premature Rupture
        20) C0015958 - Fetofetal Transfusion
        21) C0015959 - Fetomaternal Transfusion
        22) C0017506 - Gestosis, EPH
        23) C0018811 - Heart Rate, Fetal
        24) C0019343 - Herpes Gestationis
        25) C0020217 - Hydatidiform Mole
        26) C0020224 - Polyhydramnios
        27) C0020450 - Hyperemesis Gravidarum
        28) C0022864 - Labor
        29) C0022865 - Labor Complications
        30) C0022868 - Labor Onset
        31) C0022869 - Labor Presentation
        32) C0022871 - Labor Stage, First
        33) C0022872 - Labor Stage, Second
        34) C0022873 - Labor Stage, Third
        35) C0022876 - Labor, Premature
```

```
Pregnant (continued):
        36) C0024929 - Maternal-Fetal Exchange
        37) C0030612 - Parturient Paresis
        38) C0032044 - Placenta Accreta
        39) C0032045 - Placenta Diseases
        40) C0032046 - Placenta Praevia
        41) C0032051 - Placental Insufficiency
        42) C0032058 - Placentation
        43) C0032770 - Postimplantation Phase
        44) C0032797 - Postpartum Hemorrhage
        45) C0032914 - Pre-Eclampsia
        46) C0032961 - Pregnancy
        47) C0032962 - Pregnancy Complications
        48) C0032963 - Pregnancy Complications, Cardiovascular
        49) C0032964 - Pregnancy Complications, Hematologic
        50) C0032965 - Pregnancy Complications, Infectious
        51) C0032966 - Pregnancy Complications, Neoplastic
        52) C0032968 - Pregnancy in Adolescence
        53) C0032969 - Pregnancy in Diabetes
        54) C0032971 - Pregnancy Maintenance
        55) C0032972 - Pregnancy Outcome
        56) C0032975 - Pregnancy Rate
        57) C0032978 - Pregnancy Toxemias
        58) C0032984 - Pregnancy, Abdominal
        59) C0032986 - Pregnancy, Animal
        60) C0032987 - Pregnancy, Ectopic
        61) C0032989 - Pregnancy, Multiple
        62) C0032993 - Pregnancy, Prolonged
        63) C0032994 - Pregnancy, Tubal
        64) C0032995 - Pregnancy, Unwanted
        65) C0033022 - Preimplantation Phase
        66) C0033054 - Prenatal Exposure Delayed Effects
        67) C0038822 - Superfetation
        68) C0040345 - Uterine Monitoring
        69) C0040348 - Tocolysis
        70) C0040862 - Trial of Labor
        71) C0041182 - Trophoblastic Neoplasms
        72) C0042130 - Uterine Contraction
        73) C0042135 - Uterine Inertia
        74) C0042143 - Uterine Rupture
        75) C0079924 - Oligohydramnios
        76) C0080265 - Ultrasonography, Prenatal
        77) C0085207 - Diabetes, Gestational
        78) C0085547 - Phenylketonuria, Maternal
        79) C0162494 - Pregnancy Complications, Parasitic
        80) C0162739 - HELLP Syndrome
        81) C0206666 - Trophoblastic Tumor, Placental Site
        82) C0242622 - Uteroplacental Circulation
        83) C0242669 - Placenta, Retained
        84) C0242786 - Pregnancy, High-Risk
        85) C0242836 - Pregnancy Reduction, Multifetal
        86) C0342008 - Amniotic fluid pulmonary embolism
        87) C0600107 - Incomplete legal abortion
        88) C0600454 - Cervical Ripening

Pregnant Trees (9):
        G08.520.769.326.200
        G08.520.769.362
        G08.520.840
        C13.703.039
        C13.703.590
        C13.703.634
        C13.703.733
        C13.703.799
        G09.330.612.509.430
```

```
Sheep (5):
        1) C0027345 - Nairobi Sheep Disease
        2) C0032306 - Pneumonia, Progressive Interstitial, of Sheep
        3) C0034049 - Pulmonary Adenomatosis, Ovine
        4) C0036946 - Sheep Diseases
        5) C0038981 - Swayback

-----------------------------------

Swine (13):
        1) C0001752 - African Swine Fever
        2) C0013605 - Edema Disease of Swine
        3) C0014073 - Encephalomyelitis, Enzootic Porcine
        4) C0014521 - Epidermitis, Exudative, of Swine
        5) C0017162 - Gastroenteritis, Transmissible, of Swine
        6) C0019841 - Hog Cholera
        7) C0039006 - Swine Diseases
        8) C0039007 - Swine Erysipelas
        9) C0039010 - Swine Vesicular Disease
       10) C0039011 - Swine, Miniature
       11) C0042584 - Vesicular Exanthema of Swine
       12) C0242598 - LLC-PK1 Cells
       13) C0376538 - Porcine Reproductive and Respiratory Syndrome

-----------------------------------

United States (54):
        1) C0002455 - American Cancer Society
        2) C0002456 - American Dental Association
        3) C0002458 - American Heart Association
        4) C0002459 - American Hospital Association
        5) C0002461 - American Medical Association
        6) C0002463 - American Nurses' Association
        7) C0007670 - Centers for Disease Control and Prevention (U.S.)
        8) C0009434 - Commission on Professional and Hospital Activities
        9) C0018727 - Health Planning
       10) C0018763 - Health Systems Agencies
       11) C0018764 - Health Systems Plans
       12) C0020007 - Hospitals, Federal
       13) C0020012 - Hospitals, Military
       14) C0021621 - Institute of Medicine (U.S.)
       15) C0022405 - Joint Commission on Accreditation of Healthcare Organizations
       16) C0025071 - Medicaid
       17) C0025140 - MEDLARS
       18) C0025141 - MEDLINE
       19) C0027446 - National Academy of Sciences (U.S.)
       20) C0027447 - National Center for Health Care Technology
       21) C0027450 - National Center for Health Statistics (U.S.)
       22) C0027454 - National Health Insurance, United States
       23) C0027456 - National Health Planning Information Center
       24) C0027463 - National Institute for Occupational Safety and Health
       25) C0027466 - National Institute of Mental Health (U.S.)
       26) C0027468 - National Institutes of Health (U.S.)
       27) C0027470 - National Library of Medicine (U.S.)
       28) C0031826 - Physician Payment Review Commission
       29) C0033518 - Prospective Payment Assessment Commission
       30) C0038192 - State Health Planning and Development Agencies
       31) C0038194 - State Health Plans
       32) C0041704 - United States Substance Abuse and Mental Health Services Administration
       33) C0041711 - United States Dept. of Health and Human Services
       34) C0041712 - United States Environmental Protection Agency
       35) C0041713 - United States Federal Trade Commission
       36) C0041714 - United States Food and Drug Administration
       37) C0041718 - United States Health Care Financing Administration
       38) C0041720 - United States Health Resources and Services Administration
       39) C0041731 - United States Occupational Safety and Health Administration
       40) C0041732 - United States Office of Economic Opportunity
       41) C0041733 - United States Office of Technology Assessment
```

```
United States (continued):
        42) C0041734 - United States Public Health Service
        43) C0041735 - United States Department of Veterans Affairs
        44) C0078936 - American Speech-Language-Hearing Association
        45) C0080268 - United States Agency for Health Care Policy and Research
        46) C0085141 - United States Indian Health Service
        47) C0085291 - National Practitioner Data Bank
        48) C0085410 - United States Department of Agriculture
        49) C0206601 - United States Office of Research Integrity
        50) C0242776 - United States National Aeronautics and Space Administration
        51) C0282438 - Consensus Development Conferences, NIH
        52) C0282680 - United States Social Security Administration
        53) C0376631 - Employee Retirement Income Security Act
        54) C0600418 - Patient Self-Determination Act

United States Trees (4):
        N03.540.427.300
        N03.540.452.508
        N03.540.427.550
        N03.219.521.576.343.840


-----------------------------------

Mice (28):
         1) C0022827 - L Cells (Cell Line)
         2) C0025917 - Mice, Inbred A
         3) C0025918 - Mice, Inbred AKR
         4) C0025919 - Mice, Inbred BALB C
         5) C0025920 - Mice, Inbred C3H
         6) C0025921 - Mice, Inbred C57BL
         7) C0025922 - Mice, Inbred CBA
         8) C0025923 - Mice, Inbred DBA
         9) C0025924 - Mice, Inbred HRS
        10) C0025925 - Mice, Inbred ICR
        11) C0025926 - Mice, Inbred NZB
        12) C0025927 - Mice, Inbred Strains
        13) C0025928 - Mice, Jimpy
        14) C0025930 - Mice, Mutant Strains
        15) C0025931 - Mice, Neurologic Mutants
        16) C0025932 - Mice, Nude
        17) C0025933 - Mice, Obese
        18) C0025934 - Mice, Quaking
        19) C0025936 - Mice, Transgenic
        20) C0085087 - 3T3 Cells
        21) C0085112 - Mice, SCID
        22) C0085243 - Mice, Inbred NOD
        23) C0206535 - Mice, Inbred mdx
        24) C0206610 - Mice, Inbred CFTR
        25) C0206745 - Mice, Knockout
        26) C0242854 - Mice, Inbred SENCAR
        27) C0376632 - Mice, Inbred MRL lpr
        28) C0600530 - Mice, Congenic

Mice Trees (2):
        B02.649.865.635.500.380
        B02.649.865.635.500.440
```

```
Rats (22):
        1) C0034694 - Rats, Brattleboro
        2) C0034696 - Rats, Gunn
        3) C0034698 - Rats, Inbred ACI
        4) C0034699 - Rats, Inbred BB
        5) C0034700 - Rats, Inbred BN
        6) C0034701 - Rats, Inbred BUF
        7) C0034703 - Rats, Inbred F344
        8) C0034704 - Rats, Inbred Lew
        9) C0034705 - Rats, Inbred SHR
       10) C0034706 - Rats, Inbred Strains
       11) C0034707 - Rats, Inbred WF
       12) C0034709 - Rats, Inbred WKY
       13) C0034711 - Rats, Mutant Strains
       14) C0034713 - Rats, Nude
       15) C0034715 - Rats, Sprague-Dawley
       16) C0034716 - Rats, Wistar
       17) C0034719 - Rats, Zucker
       18) C0085262 - PC12 Cells
       19) C0324537 - Rats, Long-Evans
       20) C0600533 - Rats, Inbred Dahl
       21) C0600547 - Rats, Inbred LEC
       22) C0600548 - Rats, Inbred OLETF

Rat Trees (2):
        B02.649.865.635.560.500
        B02.649.865.635.560.610

----------------------------------

Hamster Trees (1):
        B02.649.865.635.325
```

Appendix D – CheckTag Lookup & Substitution List

adolescent(s)	ADOLESCENCE
adult(s)	ADULT
baby	INFANT
babies	INFANT
bovine(s)	CATTLE
boy(s)	MALE
canine(s)	DOGS
cat(s)	CATS
cattle(s)	CATTLE
child(s)	CHILD
children(s)	CHILD
dog(s)	DOGS
feline(s)	CATS
female(s)	FEMALE
girl(s)	FEMALE
hamster(s)	HAMSTERS
human(s)	HUMAN
infant(s)	INFANT
male(s)	MALE
man	MALE
men	MALE
mice	MICE
mouse	MICE
newborn(s)	INFANT, NEWBORN
paediatric(s)	CHILD
patient(s)	HUMAN
pediatric(s)	CHILD
pregnancy	PREGNANCY
pregnant	PREGNANCY
rabbit(s)	RABBITS
rat(s)	RATS
teenage	ADOLESCENCE
teenaged	ADOLESCENCE
teenager(s)	ADOLESCENCE
woman	FEMALE
women	FEMALE

Appendix E – Geographics Lookup & Substitution List

Dhaka	Bangladesh
Bangkok	Thailand
Chiang Rai	Thailand
Cape Town	South Africa
Carletonville	South Africa
Durban	South Africa
Gauteng	South Africa
Johannesburg	South Africa
Kwazulu-Natal	South Africa
Lusaka	Zambia
Ndola	Zambia
Harare	Zimbabwe
Yaounde	Cameroon
Bangui	Central African Republic
Katanga	Democratic Republic of Congo
Kinshasa	Democratic Republic of Congo
Mombasa	Kenya
Nairobi	Kenya
Nyanza Province	Kenya
Dar es Salaam	Tanzania
Kagera	Tanzania
Mwanza	Tanzania
Rakai	Uganda
Kampala	Uganda
Kigali	Rwanda
Abidjan	Cote d'Ivoire
Bouake	Cote d'Ivoire
Conakry	Guinea
Lagos	Nigeria
Dakar	Senegal
Cotonou	Benin
Guangxi	China
Shanghai	China
Yunnan	China
Manila	Philippines
Bangalore	India
Calcutta	India
Chennai	India
Karnataka	India
Madurai	India
Madras	India
Maharashtra	India
Manipur	India
Pune	India
Rajasthan	India

Appendix F – MH/SH Lookup & Substitution List

MH	SH
Abnormalities	abnormalities
Allergy and Immunology	immunology
Blood	blood
Blood Circulation	blood supply
Cerebrospinal Fluid	cerebrospinal fluid
Chemistry	chemistry
Chemistry, Analytical	analysis
Classification	classification
Cytology	cytology
Deficiency Diseases	deficiency
Diagnosis	diagnosis
Diet Therapy	diet therapy
Disease Transmission	transmission
Drug Therapy	drug therapy
Economics	economics
Education	education
Embryology	embryology
Enzymes	enzymology
Epidemiology	epidemiology
Equipment and Supplies	instrumentation
Ethnology.	ethnology
Genetics	genetics
Health Manpower	manpower
History	history
Metabolism	metabolism
Methods	methods
Microbiology	microbiology
Mortality	mortality
Neoplasm Metastasis	secondary
Nursing	nursing
Nursing Care	nursing
Organization and Administration	organization & administration
Parasitology	parasitology
Pathology	pathology
Pharmacokinetics	pharmacokinetics
Pharmacology	pharmacology
Physiology	physiology
Poisoning	poisoning
Preventive Medicine	prevention & control
Psychology	psychology
Radiation Effects	radiation effects
Radiography	radiography
Radionuclide Imaging	radionuclide imaging
Radiotherapy	radiotherapy
Rehabilitation	rehabilitation
Statistics	statistics & numerical data
Surgery	surgery
Surgical Procedures, Operative	surgery
Therapeutics	therapy
Toxicology	toxicity
Transplantation	transplantation
Ultrasonography	ultrasonography
Urine	urine
Veterinary Medicine	veterinary
Virology	virology

Appendix G – MH Exclusion List

MeSH Term	MeSH Term
Age Groups	Neoplasms by Histologic Type
Algae and Fungi	Neoplasms by Site
Amino Acid Sequence	Neurobehavioral Manifestations
Amino Acids, Peptides, and Proteins	Neurologic Manifestations
Anesthesia and Analgesia	Neuromuscular Manifestations
Anti-Allergic and Respiratory System Agents	Neurotransmitters and Neurotransmitter Agents
Anti-Inflammatory Agents, Antirheumatic Agents, and Inflammation Mediators	Nevi and Melanomas
Antineoplastic and Immunosuppressive Agents	Nucleic Acids, Nucleotides, and Nucleosides
Bacterial Infections and Mycoses	Nutritional and Metabolic Diseases
Base Sequence	Occupational Groups
Behavior and Behavior Mechanisms	Oral Manifestations
Behavioral Disciplines and Activities	Pathologic Processes
Biochemical Phenomena, Metabolism, and Nutrition	Pathological Conditions, Signs and Symptoms
Biological Phenomena, Cell Phenomena, and Immunity	Persons
Body Regions	Plant Families and Groups
Body Temperature Changes	Population Characteristics
Body Weight Changes	Psychological Phenomena and Processes
Carbohydrate Sequence	Rehabilitation of Speech and Language Disorders
Carbohydrates and Hypoglycemic Agents	Reproductive and Urinary Physiology
Cell Division Phases	Schizophrenia and Disorders with Psychotic Features
Chemical Actions	Signs and Symptoms
Chemical Actions and Uses	Signs and Symptoms, Digestive
Chemical and Pharmacologic Phenomena	Signs and Symptoms, Respiratory
Circulatory and Respiratory Physiology	Skin and Connective Tissue Diseases
Congenital, Hereditary, and Neonatal Diseases and Abnormalities	Skin Manifestations
Digestive, Oral, and Skin Physiology	Technology, Industry, and Agriculture
Disease Attributes	Tissues
Disorders of Environmental Origin	Tooth Components
Embolism and Thrombosis	Urologic and Male Genital Diseases
Environment and Public Health	Urological Manifestations
Environmental Pollutants, Noxae, and Pesticides	Uses of Chemicals and Drugs
Enzymes, Coenzymes, and Enzyme Inhibitors	Vertebrate Viruses
Epidemiologic Study Characteristics	
Eye Manifestations	
Female Genital Diseases and Pregnancy Complications	
Fluids and Secretions	
Food and Beverages	
Genes	
Geographic Locations	
Growth and Embryonic Development	
Growth Substances, Pigments, and Vitamins	
Health Care Economics and Organizations	
Health Care Evaluation Mechanisms	
Health Care Facilities, Manpower, and Services	
Health Care Quality, Access, and Evaluation	
Hemic and Immune Systems	
Hormones, Hormone Substitutes, and Hormone Antagonists	
Immunologic and Biological Factors	
Investigative Techniques	
Lipids and Antilipemic Agents	
Microbiologic Phenomena	
Molecular Sequence Data	
Musculoskeletal, Neural, and Ocular Physiology	